Extending the square root method to account for additive forecast noise in ensemble methods

Extending the Square Root Method to Account for Additive Forecast Noise in Ensemble Methods

PATRICK NIMA RAANES

Nansen Environmental and Remote Sensing Center, Bergen, Norway, and Mathematical Institute, University of Oxford, Oxford, United Kingdom

ALBERTO CARRASSI AND LAURENT BERTINO

Nansen Environmental and Remote Sensing Center, Bergen, Norway

(Manuscript received 7 November 2014, in final form 26 May 2015)

ABSTRACT

A square root approach is considered for the problem of accounting for model noise in the forecast step of the ensemble Kalman filter (EnKF) and related algorithms. The primary aim is to replace the method of simulated, pseudo-random additive so as to eliminate the associated sampling errors. The core method is based on the analysis step of ensemble square root filters, and consists in the deterministic computation of a transform matrix. The theoretical advantages regarding dynamical consistency are surveyed, applying equally well to the square root method in the analysis step. A fundamental problem due to the limited size of the ensemble subspace is discussed, and novel solutions that complement the core method are suggested and studied. Benchmarks from twin experiments with simple, low-order dynamics indicate improved performance over standard approaches such as additive, simulated noise, and multiplicative inflation.

1. Introduction

The ensemble Kalman filter (EnKF) is a popular method for doing data assimilation (DA) in the geosciences. This study is concerned with the treatment of model noise in the EnKF forecast step.

a. Relevance and scope

While uncertainty quantification is an important end product of any estimation procedure, it is paramount in DA because of the sequentiality and the need to correctly weight the observations at the next time step. The two main sources of uncertainty in a forecast are the initial conditions and model error (Slingo and Palmer 2011). Accounting for model error is therefore essential in DA.

Model error, the discrepancy between nature and computational model, can be due to incomplete understanding, linearization, truncation, subgrid-scale processes, and numerical imprecision (Nicolis 2004; Li et al. 2009). For the purposes of DA, however, model error is frequently described as a stochastic, additive, stationary, zero-centered, spatially correlated, Gaussian white noise process. This is highly unrealistic, yet defensible in view of the multitude of unknown error sources, the central limit theorem, and tractability (Jazwinski 1970, section 3.8). Another issue is that the size and complexity of geoscientific models makes it infeasible to estimate the model error statistics to a high degree of detail and accuracy, necessitating further reduction of its parameterizations (Dee 1995).

The model error in this study adheres to all of the above assumptions. This, however, renders it indistinguishable from a noise process, even from our omniscient point of view. Thus, this study effectively also pertains to natural noises not generally classified as model error, such as inherent stochasticity (e.g., quantum mechanics) and stochastic, external forcings (e.g., cosmic microwave radiation). Therefore, while model error remains the primary motivation, model *noise* is henceforth the designation most used. It is left to future studies to recuperate more generality by scaling back on the assumptions.

Several studies in the literature are concerned with the estimation of model error, as well as its treatment

Corresponding author address: Patrick Nima Raanes, Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Rd., Oxford OX2 6GG, United Kingdom.
E-mail: patrick.raanes@maths.ox.ac.uk

DOI: 10.1175/MWR-D-14-00375.1

in a DA scheme (Daley 1992; Zupanski and Zupanski 2006; Mitchell and Carrassi 2015). The scope of this study is more restricted, addressing the treatment only. To that end, it is functional to assume that the noise statistics, namely, the mean and covariance, are perfectly known. This unrealistic assumption is therefore made, allowing us to focus solely on the problem of incorporating or *accounting for* model noise in the EnKF.

b. Model noise treatment in the EnKF

From its inception, the EnKF has explicitly considered model noise and accounted for it in a Monte Carlo way: adding simulated, pseudorandom noise to the state realizations (Evensen 1994). A popular alternative technique is multiplicative inflation, where the spread of the ensemble is increased by some "inflation factor." Several comparisons of these techniques exist in the literature (e.g., Hamill and Whitaker 2005; Whitaker et al. 2008; Deng et al. 2011).

Quite frequently, however, model noise is not explicitly accounted for, but treated simultaneously with other system errors, notably sampling error and errors in the specification of the noise statistics (Whitaker et al. 2004; Hunt et al. 2004; Houtekamer et al. 2005; Anderson 2009). This is because (i) inflation can also be used to compensate for these system errors and (ii) tuning separate inflation factors seems wasteful or even infeasible. Nevertheless, even in realistic settings, it can be rewarding to treat model error explicitly. For example, Whitaker and Hamill (2012) show evidence that, in the presence of multiple sources of error, a tuned combination of a multiplicative technique and additive noise is superior to either technique used alone.

Section 5 discusses the EnKF model noise incorporation techniques most relevant to this manuscript. However, the scope of this manuscript is not to provide a full comparison of all of the alternatives under all relevant circumstances, but to focus on the square root approach. Techniques not considered any further here include using more complicated stochastic parameterizations (Arnold et al. 2013; Berry and Harlim 2014), physics-based forcings such as stochastic kinetic energy backscatter (Shutts 2005), relaxation (Zhang et al. 2004), and boundary condition forcings.

c. Framework

Suppose the state and observation, $\mathbf{x}^t \in \mathbb{R}^m$ and $\mathbf{y}^t \in \mathbb{R}^p$, respectively, are generated by

$$\mathbf{x}^{t+1} = f(\mathbf{x}^t) + \mathbf{q}^t, \quad t = 0, 1, \ldots, \tag{1}$$

$$\mathbf{y}^t = \mathbf{H}\mathbf{x}^t + \mathbf{r}^t, \quad t = 1, 2, \ldots, \tag{2}$$

where the Gaussian white noise processes $\{\mathbf{q}^t \mid t = 0, 1, \ldots\}$ and $\{\mathbf{r}^t \mid t = 1, 2, \ldots\}$, and the initial condition, $\mathbf{x}^0$, are specified by

$$\mathbf{q}^t \sim \mathcal{N}(0, \mathbf{Q}), \quad \mathbf{r}^t \sim \mathcal{N}(0, \mathbf{R}), \quad \mathbf{x}^0 \sim \mathcal{N}(\boldsymbol{\mu}^0, \mathbf{P}^0). \tag{3}$$

The observation operator, $\mathbf{H} \in \mathbb{R}^{p \times m}$, has been assumed linear because that is how it will effectively be treated anyway [e.g., through the augmentation trick of Anderson (2001)]. The parameter $\boldsymbol{\mu}^0 \in \mathbb{R}^m$ is assumed known, as are the symmetric, positive-definite (SPD) covariance matrices $\mathbf{P}^0$, $\mathbf{Q} \in \mathbb{R}^{m^2}$, and $\mathbf{R} \in \mathbb{R}^{p^2}$. Generalization to time-dependent $\mathbf{Q}$, $\mathbf{R}$, f, and $\mathbf{H}$ is straightforward.

Consider $p(\mathbf{x}^t \mid \mathbf{y}^{1:t})$, the Bayesian probability distribution of $\mathbf{x}^t$ conditioned on all of the previous observations $\mathbf{y}^{1:t}$, where the colon indicates an integer sequence. The recursive filtering process is usually broken into two steps: the forecast step, whose output is denoted by the superscript f, and the analysis step, whose output is denoted using the superscript a. Accordingly, the first and second moments of the distributions are denoted as

$$\mathbf{x}^f = \mathbb{E}(\mathbf{x}^t \mid \mathbf{y}^{1:t-1}), \quad \mathbf{P}^f = \mathrm{Var}(\mathbf{x}^t \mid \mathbf{y}^{1:t-1}), \tag{4}$$

$$\mathbf{x}^a = \mathbb{E}(\mathbf{x}^t \mid \mathbf{y}^{1:t}), \quad \mathbf{P}^a = \mathrm{Var}(\mathbf{x}^t \mid \mathbf{y}^{1:t}), \tag{5}$$

where $\mathbb{E}(\,)$ and $\mathrm{Var}(\,)$ are the (multivariate) expectation and variance operators, respectively. In the linear-Gaussian case, these characterize $p(\mathbf{x}^t \mid \mathbf{y}^{1:t-1})$ and $p(\mathbf{x}^t \mid \mathbf{y}^{1:t})$, and are given, recursively in time for sequentially increasing indices, t, by the Kalman filter equations.

The EnKF is an algorithm to approximately sample ensembles, $\mathbf{x}_{1:N} = \{\mathbf{x}_n \mid n = 1{:}N\}$, from these distributions. Note that the positive integer N is used to denote ensemble size, while m and p have been used to denote state and observation vector lengths. For convenience, all of the state realizations are assembled into the "ensemble matrix":

$$\mathbf{E} = [\mathbf{x}_1, \quad \ldots, \quad \mathbf{x}_n, \quad \ldots, \quad \mathbf{x}_N]. \tag{6}$$

A related matrix is that of the "anomalies":

$$\mathbf{A} = \mathbf{E}(\mathbf{I}_N - \boldsymbol{\Pi}_{\mathbb{1}}) = \mathbf{E}(\mathbf{I}_N - \mathbb{1}\mathbb{1}^{\mathrm{T}}/N), \tag{7}$$

where $\mathbb{1} \in \mathbb{R}^N$ is the column vector of ones, $\mathbb{1}^{\mathrm{T}}$ is its transpose, and the matrix $\mathbf{I}_N$ is the $N \times N$ identity. The conventional estimators serve as ensemble counterparts to the exact first- and second-order moments of Eqs. (4) and (5):

$$\overline{\mathbf{x}}^f = \frac{1}{N}\mathbf{E}^f\mathbb{1}, \quad \overline{\mathbf{P}}^f = \frac{1}{N-1}\mathbf{A}^f\mathbf{A}^{f\mathrm{T}}, \tag{8}$$

$$\overline{\mathbf{x}}^a = \frac{1}{N}\mathbf{E}^a\mathbb{1}, \quad \overline{\mathbf{P}}^a = \frac{1}{N-1}\mathbf{A}^a\mathbf{A}^{a\mathrm{T}}, \tag{9}$$

where, again, the superscripts indicate the conditioning. Furthermore, $\mathbf{A}$ (without any superscript) is henceforth used to refer to the anomalies at an intermediate stage in the forecast step, before model noise incorporation. In summary, the superscript usage of the EnKF cycle is illustrated by

$$\overbrace{\mathbf{A}^a \xrightarrow[\text{Eq. (28)}]{\text{Model integration,}} \mathbf{A} \xrightarrow[\text{incorporation}]{\text{Model noise}} \mathbf{A}^f}^{\text{Forecast step}} \underbrace{\xrightarrow[\text{Eqs. (17),(21)}]{\text{Analysis}} \mathbf{A}^a}_{\text{Analysis step}} .$$

Although the first $\mathbf{A}^a$ of the diagram is associated with the time step before that of $\mathbf{A}$, $\mathbf{A}^f$, and the latter $\mathbf{A}^a$, this ambiguity becomes moot by focusing on the analysis step and the forecast step separately.

d. Layout

The proposed methods to account for model noise builds on the square root method of the analysis step, which is described in section 2. The core of the proposed methods is then set forth in section 3. Properties of both methods are analyzed in section 4. Alternative techniques, against which the proposed method is compared, are outlined in section 5. Based on these alternatives, section 6 introduces methods to account for the residual noise resulting from the core method. It, therefore, connects to, and completes, section 3. The setup and results of numerical experiments are given in sections 7 and 8. A summary is provided, along with a final discussion, in section 9. The appendixes provide additional details on the properties of the proposed square root methods.

2. The square root method in the analysis step

Before introducing the square root method for the EnKF forecast step, which accounts for model noise, we here briefly discuss the square root method in the analysis step.

a. Motivation

It is desirable that $\overline{\mathbf{P}}^{a/f} = \mathbf{P}^{a/f}$ and $\overline{\mathbf{x}}^{a/f} = \mathbf{x}^{a/f}$ throughout the DA process. This means that the Kalman filter equations, with the ensemble estimates swapped in,

$$\overline{\mathbf{K}} = \overline{\mathbf{P}}^f \mathbf{H}^{\mathrm{T}} (\mathbf{H}\overline{\mathbf{P}}^f \mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{10}$$

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \overline{\mathbf{K}}[\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^f], \tag{11}$$

$$\overline{\mathbf{P}}^a = [\mathbf{I}_m - \overline{\mathbf{K}}\mathbf{H}]\overline{\mathbf{P}}^f, \tag{12}$$

should be satisfied by $\mathbf{E}^a$ from the analysis update.

Let $\mathbf{D}_{\text{obs}} \in \mathbb{R}^{p\times N}$ be a matrix whose columns are drawn independently from $\mathcal{N}(0, \mathbf{R})$. Unfortunately, the perturbed observations analysis update (Burgers et al. 1998),

$$\mathbf{E}^a = \mathbf{E}^f + \overline{\mathbf{K}}\{\mathbf{y}\mathbb{1}^{\mathrm{T}} + \mathbf{D}_{\text{obs}} - \mathbf{H}\mathbf{E}^f\}, \tag{13}$$

only yields the intended covariance, Eq. (12), on average:

$$\mathbb{E}(\overline{\mathbf{P}}^a) = [\mathbf{I}_m - \overline{\mathbf{K}}\mathbf{H}]\overline{\mathbf{P}}^f, \tag{14}$$

where the expectation, $\mathbb{E}$, is taken with respect to $\mathbf{D}_{\text{obs}}$.

b. Method

On the other hand, the square root analysis update satisfies Eq. (12) exactly. Originally introduced to the EnKF by Bishop et al. (2001), the square root analysis approach was soon connected to classic square root Kalman filters (Tippett et al. 2003). But while the primary intention of classic square root Kalman filters was to improve on the numerical stability of the Kalman filter (Anderson and Moore 1979), the main purpose of the square root EnKF was rather to eliminate the stochasticity and the accompanying sampling errors of the perturbed-observations analysis update in Eq. (13).

Assume that $p \leq m$, or that $\mathbf{R}$ is diagonal, or that $\mathbf{R}^{-1/2}$ is readily computed. Then, both for notational and computational (Hunt et al. 2007) simplicity, let

$$\mathbf{S} = \mathbf{R}^{-1/2}(\mathbf{H}\mathbf{A}^f)/\sqrt{N-1} \quad \in \mathbb{R}^{p\times N}, \tag{15}$$

$$\mathbf{s} = \mathbf{R}^{-1/2}[\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^f]/\sqrt{N-1} \in \mathbb{R}^p, \tag{16}$$

denote the "normalized" anomalies and mean innovation of the ensemble of observations. Recalling Eq. (9) it can then be shown that Eqs. (10)–(12) are satisfied if

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{A}^f \mathbf{G}^a \mathbf{S}^{\mathrm{T}} \mathbf{s}, \tag{17}$$

$$\mathbf{A}^a \mathbf{A}^{a\mathrm{T}} = \mathbf{A}^f \mathbf{G}^a \mathbf{A}^{f\mathrm{T}}, \tag{18}$$

where the two forms of $\mathbf{G}^a$,

$$\mathbf{G}^a = \mathbf{I}_N - \mathbf{S}^{\mathrm{T}}(\mathbf{S}\mathbf{S}^{\mathrm{T}} + \mathbf{I}_p)^{-1}\mathbf{S} \tag{19}$$

$$= (\mathbf{S}^{\mathrm{T}}\mathbf{S} + \mathbf{I}_N)^{-1}, \tag{20}$$

are linked through the Woodbury identity (e.g., Wunsch 2006). Therefore, if $\mathbf{A}^a$ is computed by

$$\mathbf{A}^a = \mathbf{A}^f \mathbf{T}^a, \tag{21}$$

with $\mathbf{T}^a$ being a matrix square root of $\mathbf{G}^a$, then $\mathbf{A}^a$ satisfies Eq. (12) exactly. Moreover, "square root update" is henceforth the term used to refer to any update of the anomalies through the right multiplication of a transform

matrix, as in Eq. (21). The ensemble is obtained by recombining the anomalies and the mean:

$$\mathbf{E}^a = \bar{\mathbf{x}}^a \mathbb{1}^{\mathrm{T}} + \mathbf{A}^a . \tag{22}$$

c. The symmetric square root

Equation (20) implies that $\mathbf{G}^a$ is SPD. The matrix $\mathbf{T}^a$ is a square root of $\mathbf{G}^a$ if it satisfies

$$\mathbf{G}^a = \mathbf{T}^a \mathbf{T}^{a\mathrm{T}} . \tag{23}$$

However, by substitution into Eq. (23) it is clear that $\mathbf{T}^a\mathbf{\Omega}$ is also a square root of $\mathbf{G}^a$, for any orthogonal matrix $\mathbf{\Omega}$. There are, therefore, infinitely many square roots. Nevertheless, some have properties that make them unique. For example, the Cholesky factor is unique as the only triangular square root with positive diagonal entries.

Here, however, the square root of most interest is the symmetric one, $\mathbf{T}^a_s = \mathbf{V}\mathbf{\Sigma}^{1/2}\mathbf{V}^{\mathrm{T}}$. Here, $\mathbf{V}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} = \mathbf{G}^a$ is an eigendecomposition of $\mathbf{G}^a$, and $\mathbf{\Sigma}^{1/2}$ is defined as the entry-wise *positive* square root of $\mathbf{\Sigma}$ (Horn and Johnson 2013, Theorem 7.2.6). Its existence follows from the spectral theorem, and its uniqueness from that of the eigendecomposition. Note its distinction by the s subscript.

It has gradually been discovered that the symmetric square root choice has several advantageous properties for its use in Eq. (21), one of which is that the it does not affect the ensemble mean (e.g., Wang and Bishop 2003; Evensen 2009), which is updated by Eq. (17) apart from the anomalies. Further advantages are surveyed in section 4, providing strong justification for choosing the symmetric square root, and strong motivation to extend the square root approach to the forecast step.

3. The square root method in the forecast step

Section 2 reviewed the square root update method for the analysis step of the EnKF. In view of its improvements over the Monte Carlo method, it is expected that a similar scheme for incorporating the model noise into the forecast ensemble, $\mathbf{E}^f$, would be beneficial. Section 3b derives such a scheme: SQRT-CORE. First, however, section 3a illuminates the motivation: forecast step sampling error.

a. Forecast sampling errors in the classic EnKF

Assume linear dynamics, $f: \mathbf{x} \mapsto f(\mathbf{x}) = \mathbf{F}\mathbf{x}$, for ease of illustration. The Monte Carlo simulation of Eq. (1) can be written as

$$\mathbf{E}^f = \mathbf{F}\mathbf{E}^a + \mathbf{D}, \tag{24}$$

where the columns of $\mathbf{D}$ are drawn from $\mathcal{N}(0, \mathbf{Q})$ by

$$\mathbf{D} = \mathbf{Q}^{1/2}\mathbf{\Xi}, \tag{25}$$

where $\mathbf{\Xi} = [\boldsymbol{\xi}_1, \ \ldots, \ \boldsymbol{\xi}_n, \ \ldots, \ \boldsymbol{\xi}_N]$, and each $\boldsymbol{\xi}_n$ is independently drawn from $\mathcal{N}(0, \mathbf{I}_m)$. Note that different choices of the square root, say $\mathbf{Q}^{1/2}$ and $\mathbf{Q}^{1/2}\mathbf{\Omega}$, yield equally distributed random variables, $\mathbf{Q}^{1/2}\boldsymbol{\xi}$ and $\mathbf{Q}^{1/2}\mathbf{\Omega}\boldsymbol{\xi}$. Therefore, the choice does not matter, and is left unspecified. It is typical to eliminate sampling error of the first order by centering the model noise perturbations so that $\mathbf{D}\mathbb{1} = 0$. This introduces dependence between the samples and reduces the variance. The latter is compensated for by rescaling by a factor of $\sqrt{N/(N-1)}$. The result is that

$$\overline{\mathbf{P}}^f = \mathbf{F}\overline{\mathbf{P}}^a\mathbf{F}^{\mathrm{T}} + \mathbf{Q} + (\overline{\mathbf{Q}} - \mathbf{Q}) - \frac{1}{N-1}[\mathbf{F}\mathbf{A}^a\mathbf{D}^{\mathrm{T}} + \mathbf{D}(\mathbf{F}\mathbf{A}^a)^{\mathrm{T}}], \tag{26}$$

as per Eq. (8), where $\overline{\mathbf{Q}} = (N-1)^{-1}\mathbf{D}\mathbf{D}^{\mathrm{T}}$. But, for the same reasons as for the analysis step, ideally

$$\overline{\mathbf{P}}^f = \mathbf{F}\overline{\mathbf{P}}^a\mathbf{F}^{\mathrm{T}} + \mathbf{Q}. \tag{27}$$

Thus, the second line of Eq. (26) constitutes a stochastic discrepancy from the desired relations (27).

b. The square root method for model noise: SQRT-CORE

As illustrated in section 1c, define $\mathbf{A}$ as the anomalies of the propagated ensemble before noise incorporation:

$$\mathbf{A} = f(\mathbf{E}^a)(\mathbf{I}_N - \mathbb{1}\mathbb{1}^{\mathrm{T}}/N), \tag{28}$$

where f is applied column-wise to $\mathbf{E}^a$. Then the desired relation (27) is satisfied if $\mathbf{A}^f$ satisfies

$$\mathbf{A}^f\mathbf{A}^{f^{\mathrm{T}}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} + (N-1)\mathbf{Q}. \tag{29}$$

However, $\mathbf{A}^f$ can only have N columns. Thus, the problem of finding an $\mathbf{A}^f$ that satisfies Eq. (29) is ill posed, since the right-hand side of Eq. (29) is of rank m for arbitrary, full-rank $\mathbf{Q}$, while the left-hand side is of rank N or less.

Therefore, let $\mathbf{A}^+$ be the Moore–Penrose pseudoinverse of $\mathbf{A}$, denote $\mathbf{\Pi}_{\mathbf{A}} = \mathbf{A}\mathbf{A}^+$ the orthogonal projector onto the column space of $\mathbf{A}$, and define $\hat{\mathbf{Q}} = \mathbf{\Pi}_{\mathbf{A}}\mathbf{Q}\mathbf{\Pi}_{\mathbf{A}}$ the "two-sided" projection of $\mathbf{Q}$. Note that the orthogonality of the projector, $\mathbf{\Pi}_{\mathbf{A}}$, induces its symmetry. Instead of Eq. (29), the core square root model noise incorporation method proposed here, SQRT-CORE, only aims to satisfy

$$\mathbf{A}^f \mathbf{A}^{f\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} + (N-1)\hat{\mathbf{Q}}. \tag{30}$$

By virtue of the projection, Eq. (30) can be written as

$$\mathbf{G}^f = \mathbf{I}_N + (N-1)\mathbf{A}^+ \mathbf{Q} (\mathbf{A}^+)^{\mathrm{T}}, \tag{31}$$

$$\mathbf{A}^f \mathbf{A}^{f\mathrm{T}} = \mathbf{A}\mathbf{G}^f \mathbf{A}^{\mathrm{T}}. \tag{32}$$

Thus, with $\mathbf{T}^f$ being a square root of $\mathbf{G}^f$, the update

$$\mathbf{A}^f = \mathbf{A}\mathbf{T}^f \tag{33}$$

accounts for the component of the noise quantified by $\hat{\mathbf{Q}}$. The difference between the right-hand sides of Eqs. (29) and (30), $(N-1)[\mathbf{Q} - \hat{\mathbf{Q}}]$, is henceforth referred to as the "residual noise" covariance matrix. Accounting for it is not trivial. This discussion is resumed in section 6.

As for the analysis step, we choose to use the symmetric square root, $\mathbf{T}^f_s$, of $\mathbf{G}^f$. Note that *two* SVDs are required to perform this step: one to calculate $\mathbf{A}^+$ and one to calculate the symmetric square root of $\mathbf{G}^f$. Fortunately, both are relatively computationally inexpensive, needing only to calculate $N-1$ singular values and vectors. For later use, define the square root "additive equivalent":

$$\hat{\mathbf{D}} = \mathbf{A}^f - \mathbf{A} = \mathbf{A}[\mathbf{T}^f_s - \mathbf{I}_N]. \tag{34}$$

c. *Preservation of the mean*

The square root update is a deterministic scheme that satisfies the covariance update relations exactly (in the space of $\mathbf{A}$). But in updating the anomalies, the mean should remain the same. For SQRT-CORE, this can be shown to hold true in the same way as Livings et al. (2008) did for the analysis step, with the addition of Eq. (36).

Theorem 1 (mean preservation): If $\mathbf{A}^f = \mathbf{A}\mathbf{T}^f_s$, then

$$\mathbf{A}^f \mathbb{1} = 0. \tag{35}$$

That is, the symmetric square root choice for the model noise transform matrix preserves the ensemble mean.

Proof: For any matrix $\mathbf{A}$,

$$\mathbf{A}^+ = \mathbf{A}^{\mathrm{T}} (\mathbf{A}\mathbf{A}^{\mathrm{T}})^+, \tag{36}$$

(Ben-Israel and Greville 2003, section 1.6). Thus,

$$\mathbf{G}^f \mathbb{1} = \mathbb{1} + (N-1)\mathbf{A}^+ \mathbf{Q} (\mathbf{A}\mathbf{A}^{\mathrm{T}})^+ \mathbf{A}\mathbb{1} = \mathbb{1}, \tag{37}$$

as per Eq. (28). But the eigenvectors of the square of a diagonalizable matrix are the same as for the original matrix, with squared eigenvalues. Thus, Eq. (37) implies $\mathbf{A}^f \mathbb{1} = \mathbf{A}\mathbf{T}^f_s \mathbb{1} = \mathbf{A}\mathbb{1} = 0$. □ [The open square symbol (□) indicates completion of a proof.]

4. Dynamical consistency of square root updates

Many dynamical systems embody "balances" or constraints on the state space (van Leeuwen 2009). For reasons of complexity and efficiency these concerns are often not encoded in the prior (Wang et al. 2015, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*). They are, therefore, not considered by the statistical updates, resulting in state realizations that are inadmissible because of a lack of dynamical consistency or physical feasibility. Typical consequence of breaking such constraints include unbounded growth ("blow up"), exemplified by the quasigeostrophic model of Sakov and Oke (2008a), or failure of the model to converge, exemplified by reservoir simulators (Chen and Oliver 2013).

This section provides a formal review of the properties of the square root update as regards dynamical consistency, presenting theoretical support for the square root method. The discussion concerns any square root update, and is therefore relevant for the square root method in the analysis step as well as for SQRT-CORE.

a. *Affine subspace confinement*

The fact that the square root update $\mathbf{A} \mapsto \mathbf{A}\mathbf{T}$ is a right multiplication means that each column of the updated anomalies is a linear combination of the original anomalies. On the other hand, $\mathbf{T}$ itself depends on $\mathbf{A}$. In recognition of these two aspects, Evensen (2003) called such an update a "weakly nonlinear combination." However, our preference is to describe the update as confined to the affine subspace of the original ensemble, that is the affine space $\overline{\mathbf{x}} + \mathrm{span}(\mathbf{A})$.

b. *Satisfying equality constraints*

It seems reasonable to assume that the updated ensemble, being in the space of the original one, stands a fair chance of being dynamically consistent. However, if consistency can be described as equality constraints, then discussions thereof can be made much more formal and specific, as is the purpose of this subsection. In so doing, it uncovers a couple of interesting, hitherto unnoticed advantage of the symmetric square root choice.

Suppose the original ensemble, $\mathbf{x}_{1:N}$, or $\mathbf{E}$, satisfies $\mathbf{C}\mathbf{x}_n = \mathbf{d}$ for all $n = 1{:}N$, that is,

$$\mathbf{C}\mathbf{E} = \mathbf{d}\mathbb{1}^{\mathrm{T}}. \tag{38}$$

One example is conservation of mass, in which case the state, $\mathbf{x}$, would contain grid-block densities, while the constraint coefficients, $\mathbf{C}$, would be a row vector of the corresponding volumes, and $\mathbf{d}$ would be the total mass. Another example is geostrophic balance (e.g., Hoang et al. 2005), in which case $\mathbf{x}$ would hold horizontal

velocity components and sea surface heights, while $\mathbf{C}$ would concatenate the identity and a discretized horizontal differentiation operator, and $\mathbf{d}$ would be zero.

The constraints (38) should hold also after the update. Visibly, if $\mathbf{d}$ is zero, any right multiplication of $\mathbf{E}$ (i.e., any combination of its columns) will also satisfy the constraints. This provides formal justification for the proposition of Evensen (2003) that the "linearity" of the EnKF update implicitly ensures respecting linear constraints.

One can also write

$$\mathbf{C}\bar{\mathbf{x}} = \mathbf{d}, \tag{39}$$

$$\mathbf{CA} = \mathbf{0}\mathbb{1}^{\mathrm{T}}, \tag{40}$$

implying Eq. (38) provided $\mathbf{E} = \bar{\mathbf{x}}\mathbb{1}^{\mathrm{T}} + \mathbf{A}$ holds. Equations (39) and (40) show that the ensemble mean and anomalies can be thought of as particular and homogeneous solutions to the constraints. They also indicate that in a square root update, even if $\mathbf{d}$ is not zero, one only needs to ensure that the mean constraints are satisfied, because the homogeneity of Eq. (40) means that any right-multiplying update to $\mathbf{A}$ will satisfy the anomaly constraints. However, as mentioned above, unless it preserves the mean, it might perturb Eq. (39). A corollary of Theorem 1 is therefore that the symmetric choice for the square root update also satisfies inhomogeneous constraints.

Finally, in the case of nonlinear constraints, that is, $\mathscr{C}(\mathbf{x}_n) = \mathbf{d}$, truncating the Taylor expansion of $\mathscr{C}$ yields

$$\mathbf{CA} \approx [\mathbf{d} - \mathscr{C}(\bar{\mathbf{x}})]\mathbb{1}^{\mathrm{T}}, \tag{41}$$

where $\mathbf{C} = \partial\mathscr{C}/\partial\mathbf{x}(\bar{\mathbf{x}})$. Contrary to Eq. (40), the approximate constraints of Eq. (41), are not homogeneous, and therefore not satisfied by any right-multiplying update. Again, however, by Theorem 1, the symmetric square root appears an advantageous choice, because it has 1 as an eigenvector with eigenvalue 1, and therefore satisfies the (approximate) constraints.

c. *Optimality of the symmetric choice*

A number of related properties on the optimality of the symmetric square root exist scattered in the literature. However, to the best of our knowledge, these have yet to be reunited into a unified discussion. Similarly, considerations on their implications on DA have so far not been collected. These are the aims of this subsection.

Theorem 2 (Minimal ensemble displacement): Consider the ensemble anomalies $\mathbf{A}$ with ensemble covariance matrix $\bar{\mathbf{P}}$, and let $\mathbf{q}_n$ be column n of $\mathbf{D} = \mathbf{AT} - \mathbf{A}$: the displacement of the nth anomaly through a square root update. The symmetric square root, $\mathbf{T}_s$, minimizes

$$J(\mathbf{T}) = \frac{1}{N-1}\sum_n \|\mathbf{q}_n\|^2_{\bar{\mathbf{P}}}, \tag{42}$$

$$= \mathrm{trace}([\mathbf{AT} - \mathbf{A}]^{\mathrm{T}}(\mathbf{AA}^{\mathrm{T}})^{+}[\mathbf{AT} - \mathbf{A}]) \tag{43}$$

among all $\mathbf{T} \in \mathbb{R}^{N^2}$ such that $\mathbf{ATT}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = \mathbf{AGA}^{\mathrm{T}}$, for some SPD matrix $\mathbf{G}$. Equation (43) coincides with Eq. (42) if $\bar{\mathbf{P}}^{-1}$ exists, but is also valid if not.

Theorem 2 was proven by Ott et al. (2004), and later restated by Hunt et al. (2007) as the constrained optimum of the Frobenius norm of $[\mathbf{T} - \mathbf{I}_N]$. Another interesting and desirable property of the symmetric square root is the fact that the updated ensemble members are all equally likely realizations of the estimated posterior (Wang et al. 2004; McLay et al. 2008). More recently, the choice of mapping between the original and the updated ensembles has been formulated through optimal transport theory (Reich and Cotter 2013; Oliver 2014). However, the cost functions therein typically use a different weighting on the norm than $J(\mathbf{T})$, in one case yielding an optimum that is the symmetric *left*-multiplying transform matrix—not to be confused with the right-multiplying one of Theorem 2.

Theorem 2 and the related properties should benefit the performance of filters employing the square root update, whether for the analysis step, the model noise incorporation, or both. In part, this is conjectured since minimizing the displacement of an update means that the ensemble cloud should retain some of its shape, and with it higher-order, non-Gaussian information, as illustrated in Fig. 1.

A different set of reasons to expect strong performance from the symmetric square root choice is that it should promote dynamical consistency, particularly regarding inequality constraints, such as the inherent positivity of concentration variables, as well as nonlinear equality constraints, initially discussed in section 4b. In either case it stands to reason that smaller displacements are less likely to break the constraints, and therefore that their minimization should inhibit it. Additionally, it is important when using "local analysis" localization that the ensemble is updated similarly at nearby grid points. Statistically, this is ensured by employing smoothly decaying localization functions, so that $\mathbf{G}$ does not jump too much from one grid point to the next. But, as pointed out by Hunt et al. (2007), in order to translate this smoothness onto the dynamical consistency, it is also crucial that the square root is continuous in $\mathbf{G}$. Furthermore, even if $\mathbf{G}$ does jump from one grid point to the next, it still seems plausible that the minimization of displacement might restrain the creation of dynamical inconsistencies.

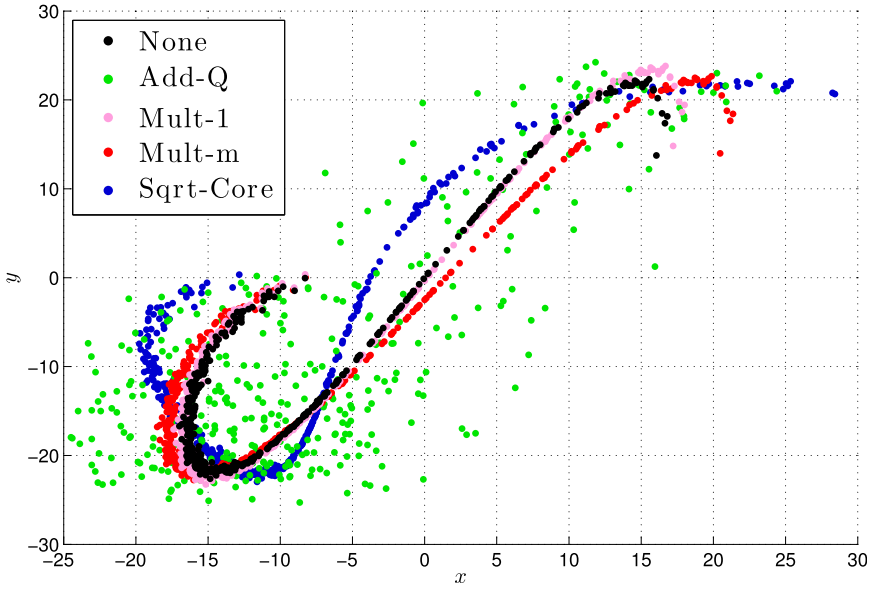

FIG. 1. Scatterplot of ensemble forecasts with the three-dimensional Lorenz-63 system (Lorenz 1963) using different schemes to account for the model noise, which is specified by $\Delta t\mathbf{Q} = \text{diag}([36.00, 3.60, 1.08])$ and makes up approximately 30% of the total spread of the updated ensembles. Each dot corresponds to the "(x, y)" coordinate of one realization among $N = 400$.

5. Alternative approaches

This section describes the model noise incorporation methods most relevant methods to this study. Table 1 summarizes the methods that will be used in numerical comparison experiments. ADD-Q is the classic method detailed in section 3a. MULT-1 and MULT-m are multiplicative inflation methods. The rightmost column relates the different methods to each other by succinctly expressing the degree to which they satisfy Eq. (29); it can also be used as a starting point for their derivation. Note that MULT-1 only satisfies one degree of freedom of Eq. (29), while MULT-m satisfies m degrees, and would therefore be expected to perform better in general. It is clear that MULT-1 and MULT-m will generally not provide an exact statistical update, no matter how big N is, while ADD-Q reproduces *all* of the moments almost surely as $N \to \infty$. By comparison, SQRT-CORE guarantees obtaining the correct first two moments for any $N > m$, but does not guarantee higher-order moments.

Using a large ensemble size, Fig. 1 illustrates the different techniques. Notably, the cloud of ADD-Q is clearly more dispersed than any of the other methods. Furthermore, in comparison to MULT-m and MULT-1, SQRT-CORE significantly skewers the distribution in order to satisfy the off-diagonal conditions.

Continuing from section 1b, the following details other pertinent alternatives, some of them sharing some similarity with the square root methods proposed here.

One alternative is to resample the ensemble fully from $\mathcal{N}(0, \mathbf{A}\mathbf{A}^{\mathrm{T}}/(N-1) + \mathbf{Q})$. However, this incurs larger sampling errors than ADD-Q, and is more likely to cause dynamical inconsistencies.

Second-order exact sampling (Pham 2001) attempts to sample noise under the restriction that all of the terms on the second line of Eq. (27) be zero. It requires a very large ensemble size ($N > 2m$), and is therefore typically not applicable, though recent work indicate that this might be circumvented (Hoteit et al. 2015).

The singular evolutive interpolated Kalman (SEIK) filter (Hoteit et al. 2002) has a slightly less primitive and

TABLE 1. Comparison of some model noise incorporation methods.

Description	Label	$\mathbf{A}^f =$	Where	Thus, satisfying
Additive, simulated noise	ADD-Q	$\mathbf{A} + \mathbf{D}$	$\mathbf{D}$ is a centered sample from $\mathcal{N}(0, \mathbf{Q})$	$\mathbb{E}_{\mathbf{D}}$(Eq. (29))
Scalar inflation	MULT-1	$\lambda\mathbf{A}$	$\lambda^2 = \text{trace}(\overline{\mathbf{P}})^{-1}\text{trace}(\overline{\mathbf{P}} + \mathbf{Q})$	trace(Eq. (29))
Multivariate inflation	MULT-m	$\mathbf{\Lambda}\mathbf{A}$	$\mathbf{\Lambda}^2 = \text{diag}(\overline{\mathbf{P}})^{-1}\text{diag}(\overline{\mathbf{P}} + \mathbf{Q})$	diag(Eq. (29))
Core square root method	SQRT-CORE	$\mathbf{A}\mathbf{T}$	$\mathbf{T} = [\mathbf{I}_N + (N-1)\mathbf{A}^{+}\mathbf{Q}\mathbf{A}^{+\mathrm{T}}]_s^{1/2}$	$\mathbf{\Pi}_{\mathbf{A}}$(Eq. (29))$\mathbf{\Pi}_{\mathbf{A}}$

intuitive formalism than the EnKF, typically working with matrices of size $m \times (N-1)$. Moreover, it does not have a separate step to deal with model noise, treating it instead implicitly, as part of the analysis step. This lack of modularity has the drawback that the frequency of model noise incorporation is not controllable: in case of multiple model integration steps between observations, the noise should be incorporated at each step in order to evolve with the dynamics; under different circumstances, skipping the treatment of noise for a few steps can be cost efficient (Evensen and van Leeuwen 1996). Nevertheless, a stand-alone model noise step can be distilled from the SEIK algorithm as a whole. Its forecast covariance matrix, $\overline{\mathbf{P}}^f$, would equal to that of SQRT-CORE: $\mathbf{\Pi}_\mathbf{A}(\overline{\mathbf{P}}+\mathbf{Q})\mathbf{\Pi}_\mathbf{A}$. However, unlike SQRT-CORE, which uses the symmetric square root, the SEIK uses random rotation matrices to update the ensemble. Also, the SEIK filter uses a "forgetting factor." Among other system errors, this is intended to account for the residual noise covariance, $[\mathbf{Q}-\hat{\mathbf{Q}}]$. This issue is discussed in the next section in relation to SQRT-CORE. As outlined in section 1b, however, this factor is not explicitly a function of $[\mathbf{Q}-\hat{\mathbf{Q}}]$; it is instead obtained from manual tuning. Moreover, it is only applied in the update of the ensemble mean.

Another method is to include only the $N-1$ largest eigenvalue components of $\overline{\mathbf{P}}+\mathbf{Q}$, as in reduced-rank square root filters (Verlaan and Heemink 1997), and some versions of the unscented Kalman filter (Chandrasekar et al. 2008). This method can be referred to as T-SVD because the update can be effectuated through a truncated SVD of $[\overline{\mathbf{P}}^{1/2}, \mathbf{Q}^{1/2}]$, where the choices of square roots do not matter. It captures more of the total variance than SQRT-CORE, but also changes the ensemble subspace. Moreover, it is not clear how to choose the updated ensemble. For example, one would suspect dynamical inconsistencies to arise from using the ordered sequence of the truncated SVD. Right multiplying by random rotation matrices, as in the SEIK, might be a good solution. Or, if computed in terms of a *left*-multiplying transform matrix, the symmetric choice is likely a good one. Building on T-SVD, the "partially orthogonal" EnKF and the complementary orthogonal subspace filter for efficient ensembles (COFFEE) algorithm (Heemink et al. 2001; Hanea et al. 2007) also recognize the issue of the residual noise. In contrasts with the treatments proposed in this study, these methods introduce a complementary ensemble to account for it.

6. Improving SQRT-CORE: Accounting for the residual noise

As explained in section 4a, SQRT-CORE can only incorporate noise components that are in the span (range) of $\mathbf{A}$. This leaves a residual noise component unaccounted for, orthogonal to the span of $\mathbf{A}$, with $[\mathbf{Q}-\hat{\mathbf{Q}}]$ posing as its covariance matrix.

First consider *why* there is no such residual of $\mathbf{R}$ for the square root methods in the analysis step: because the analysis step subtracts uncertainty, unlike the forecast step which adds it. Therefore, the presence or absence of components of $\mathbf{R}$ outside of the span of the observation ensemble makes no difference to the analysis covariance update because the ensemble effectively already assumes zero uncertainty in these directions.

In the rest of this section the question addressed is how to deal with the residual noise. It is assumed that SQRT-CORE, Eq. (33), has already been performed. The techniques proposed thus *complement* SQRT-CORE, but do not themselves possess the beneficial properties of SQRT-CORE discussed in section 4. Also, the notation of the previous section is reused. Thus, the aim of this section is to find an $\mathbf{A}^f \in \mathbb{R}^{m\times N}$ that satisfies, in some limited sense

$$\mathbf{A}^f\mathbf{A}^{f\mathrm{T}} = \mathbf{A}\mathbf{A}^\mathrm{T} + (N-1)[\mathbf{Q}-\hat{\mathbf{Q}}]. \tag{44}$$

a. Complementary, additive sampling: SQRT-ADD-Z

Let $\mathbf{Q}^{1/2}$ be *any* square root of $\mathbf{Q}$, and define

$$\hat{\mathbf{Q}}^{1/2} = \mathbf{\Pi}_\mathbf{A}\mathbf{Q}^{1/2}, \tag{45}$$

$$\mathbf{Z} = (\mathbf{I}_m - \mathbf{\Pi}_\mathbf{A})\mathbf{Q}^{1/2}, \tag{46}$$

the orthogonal projection of $\mathbf{Q}^{1/2}$ onto the column space of $\mathbf{A}$, and the complement, respectively.

A first suggestion to account for the residual noise is to use one of the techniques of section 5, with $[\mathbf{Q}-\hat{\mathbf{Q}}]$ taking the place of the full $\mathbf{Q}$ in their formulas. In particular, with ADD-Q in mind, the fact that

$$\mathbf{Q}^{1/2} = \hat{\mathbf{Q}}^{1/2} + \mathbf{Z} \tag{47}$$

motivates sampling the residual noise using $\mathbf{Z}$. That is, in addition to $\hat{\mathbf{D}}$ of SQRT-CORE, which accounts for $\hat{\mathbf{Q}}$, one also adds $\tilde{\mathbf{D}} = \mathbf{Z}\tilde{\mathbf{\Xi}}$ to the ensemble, where the columns of $\tilde{\mathbf{\Xi}}$ are drawn independently from $\mathcal{N}(0, \mathbf{I}_m)$. We call this technique SQRT-ADD-Z.

Note that $\hat{\mathbf{Q}}^{1/2}$, defined by Eq. (45), is a square root of $\hat{\mathbf{Q}}$. By contrast, multiplying Eq. (47) with its own transpose yields

$$\mathbf{Z}\mathbf{Z}^\mathrm{T} = [\mathbf{Q}-\hat{\mathbf{Q}}] - \hat{\mathbf{Q}}^{1/2}\mathbf{Z}^\mathrm{T} - \mathbf{Z}\hat{\mathbf{Q}}^{\mathrm{T}/2}, \tag{48}$$

and reveals that $\mathbf{Z}$ is not a square root of $[\mathbf{Q}-\hat{\mathbf{Q}}]$. Therefore, with expectation over $\tilde{\mathbf{\Xi}}$, SQRT-ADD-Z does not respect $\mathbb{E}$(Eq. (44)), as one would hope.

Thus, SQRT-ADD-Z has a bias equal to the cross term sum, $\hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}} + \mathbf{Z}\hat{\mathbf{Q}}^{\mathrm{T}/2} = [\mathbf{Q} - \hat{\mathbf{Q}}] - \mathbf{Z}\mathbf{Z}^{\mathrm{T}}$. Notwithstanding this problem, Corollary 1 of appendix A shows that the cross-term sum has a spectrum symmetric around 0, and thus zero trace. To some extent, this exonerates SQRT-ADD-Z, since it means that the expected total variance is unbiased.

b. *The underlying problem: Replacing a single draw with two independent draws*

Since any element of $\hat{\mathbf{Q}}$ is smaller than the corresponding element in $\mathbf{Q}$, either one of the multiplicative inflation techniques can be applied to account for $[\mathbf{Q} - \hat{\mathbf{Q}}]$ without second thoughts. Using MULT-1 would satisfy trace [Eq. (44)], while MULT-m would satisfy diag [Eq. (44)]. However, the problem highlighted for SQRT-ADD-Z is not just a technicality. In fact, as shown in section b in appendix A, $[\mathbf{Q} - \hat{\mathbf{Q}}]$ has negative eigenvalues because of the cross terms. It is therefore not a valid covariance matrix in the sense that it has no real square root: samples with covariance $[\mathbf{Q} - \hat{\mathbf{Q}}]$ will necessarily be complex numbers; this would generally be physically unrealizable and therefore inadmissible. This underlying problem seems to question the validity of the whole approach of splitting up $\mathbf{Q}$ and dealing with the parts $\hat{\mathbf{Q}}$ and $[\mathbf{Q} - \hat{\mathbf{Q}}]$ separately.

Let use emphasize the word independently, because that is, to a first approximation, what we are attempting to do: replacing a single draw from $\mathbf{Q}$ by one from $\hat{\mathbf{Q}}$ plus another, independent draw from $[\mathbf{Q} - \hat{\mathbf{Q}}]$. Rather than considering N anomalies, let us now focus on a single one, and drop the n index. Define the two random variables:

$$\mathbf{q} = \hat{\mathbf{Q}}^{1/2}\boldsymbol{\xi} + \mathbf{Z}\boldsymbol{\xi}, \tag{49}$$

$$\mathbf{q}^{\perp\perp} = \hat{\mathbf{Q}}^{1/2}\hat{\boldsymbol{\xi}} + \mathbf{Z}\tilde{\boldsymbol{\xi}}, \tag{50}$$

where $\boldsymbol{\xi}$, $\hat{\boldsymbol{\xi}}$, and $\tilde{\boldsymbol{\xi}}$ are random variables independently drawn from $\mathcal{N}(0, \mathbf{I}_m)$. By Eq. (47), and design, $\mathbf{q}$ can be identified with any of the columns of $\mathbf{D}$ of Eq. (25) and, furthermore, $\mathrm{Var}(\mathbf{q}) = \mathbf{Q}$. On the other hand, while $\mathbf{q}$ originates in a single random draw, $\mathbf{q}^{\perp\perp}$ is the sum of two independent draws.

The dependence between the terms of $\mathbf{q}$, and the lack thereof for $\mathbf{q}^{\perp\perp}$, yields the following discrepancy between the variances:

$$\mathrm{Var}(\mathbf{q}) = \hat{\mathbf{Q}} + \mathbf{Z}\mathbf{Z}^{\mathrm{T}} + \hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}} + \mathbf{Z}\hat{\mathbf{Q}}^{\mathrm{T}/2}, \tag{51}$$

$$\mathrm{Var}(\mathbf{q}^{\perp\perp}) = \hat{\mathbf{Q}} + \mathbf{Z}\mathbf{Z}^{\mathrm{T}}. \tag{52}$$

Formally, this is the same problem that was identified with Eq. (48), namely, that of finding a real square root of $[\mathbf{Q} - \hat{\mathbf{Q}}]$, or eliminating the cross terms. But Eqs. (51) and (52) show that the problem arises from the more primal problem of trying to emulate $\mathbf{q}$ by $\mathbf{q}^{\perp\perp}$. Vice versa, $\hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}} = 0$ would imply that the ostentatiously dependent terms, $\hat{\mathbf{Q}}^{1/2}\boldsymbol{\xi}$ and $\mathbf{Z}\boldsymbol{\xi}$, are independent, and thus $\mathbf{q}^{\perp\perp}$ is emulated by $\mathbf{q}$.

c. *Reintroducing dependence: SQRT-DEP*

As already noted, though, making the cross terms zero is not possible for general $\mathbf{A}$ and $\mathbf{Q}$. However, the perspective of $\mathbf{q}$ and $\mathbf{q}^{\perp\perp}$ hints at another approach: reintroducing dependence between the draws. In this section we will reintroduce dependence by making the residual sampling depend on the square root equivalent, $\hat{\mathbf{D}}$ of Eq. (34).

The trouble with the cross terms is that $\mathbf{Q}$ ''gets in the way'' between $\boldsymbol{\Pi}_{\mathbf{A}}$ and $(\mathbf{I}_m - \boldsymbol{\Pi}_{\mathbf{A}})$, whose product would otherwise be zero. Although less ambitious than emulating $\mathbf{q}$ with $\mathbf{q}^{\perp\perp}$, it is possible to emulate a single draw (e.g., $\boldsymbol{\xi}$) from $\mathcal{N}(0, \mathbf{I}_m)$ with the two independent draws of

$$\boldsymbol{\xi}^{\perp\perp} = \boldsymbol{\Pi}\hat{\boldsymbol{\xi}} + (\mathbf{I}_m - \boldsymbol{\Pi})\tilde{\boldsymbol{\xi}}, \tag{53}$$

where, as before, $\hat{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\xi}}$ are independent random variables with law $\mathcal{N}(0, \mathbf{I}_m)$, and $\boldsymbol{\Pi}$ is some orthogonal projection matrix. Then, as the cross terms cancel,

$$\boldsymbol{\Pi}\boldsymbol{\Pi}^{\mathrm{T}} + (\mathbf{I}_m - \boldsymbol{\Pi})(\mathbf{I}_m - \boldsymbol{\Pi})^{\mathrm{T}} = \mathbf{I}_m, \tag{54}$$

and thus $\mathrm{Var}(\boldsymbol{\xi}^{\perp\perp}) = \mathrm{Var}(\boldsymbol{\xi})$.

We can take advantage of this emulation possibility by choosing $\boldsymbol{\Pi}$ as the orthogonal projector onto the *rows* of $\hat{\mathbf{Q}}^{1/2}$. Instead of Eq. (49), redefine $\mathbf{q}$ as

$$\mathbf{q} = \mathbf{Q}^{1/2}\boldsymbol{\xi}^{\perp\perp}. \tag{55}$$

Then, since $\mathrm{Var}(\boldsymbol{\xi}^{\perp\perp}) = \mathbf{I}_m$,

$$\mathrm{Var}(\mathbf{q}) = \mathbf{Q}^{1/2}\mathbf{I}_m\mathbf{Q}^{\mathrm{T}/2} = \mathbf{Q}, \tag{56}$$

as desired. But also

$$\mathbf{q} = (\hat{\mathbf{Q}}^{1/2} + \mathbf{Z})[\boldsymbol{\Pi}\hat{\boldsymbol{\xi}} + (\mathbf{I}_m - \boldsymbol{\Pi})\tilde{\boldsymbol{\xi}}], \tag{57}$$

$$= \hat{\mathbf{Q}}^{1/2}\hat{\boldsymbol{\xi}} + \mathbf{Z}[\boldsymbol{\Pi}\hat{\boldsymbol{\xi}} + (\mathbf{I}_m - \boldsymbol{\Pi})\tilde{\boldsymbol{\xi}}]. \tag{58}$$

The point is that, while maintaining $\mathrm{Var}(\mathbf{q}) = \mathbf{Q}$, and despite the reintroduction of dependence between the two terms in Eq. (58), the influence of $\tilde{\boldsymbol{\xi}}$ has been confined to $\mathrm{span}(\mathbf{Z}) = \mathrm{span}(\mathbf{A})^{\perp}$. The above reflections yield the following algorithm, labeled SQRT-DEP:

1) Perform the core square root update for $\hat{\mathbf{Q}}$, Eq. (33);
2) Find $\hat{\boldsymbol{\Xi}}$ such that $\hat{\mathbf{Q}}_s^{1/2}\hat{\boldsymbol{\Xi}} = \hat{\mathbf{D}}$ of Eq. (34). Components in the kernel of $\hat{\mathbf{Q}}_s^{1/2}$ are inconsequential;
3) Sample $\tilde{\boldsymbol{\Xi}}$ by drawing each column independently from $\mathcal{N}(0, \mathbf{I}_m)$;
4) Compute the residual noise, $\tilde{\mathbf{D}}$, and add it to the ensemble anomalies:

$$\tilde{\mathbf{D}} = \mathbf{Z}[\boldsymbol{\Pi}\hat{\boldsymbol{\Xi}} + (\mathbf{I}_m - \boldsymbol{\Pi})\tilde{\boldsymbol{\Xi}}]. \tag{59}$$

Unfortunately, this algorithm requires the additional SVD of $\hat{\mathbf{Q}}^{1/2}$ in order to compute $\boldsymbol{\Pi}$ and $\hat{\boldsymbol{\Xi}}$. Also, despite the reintroduction of dependence, SQRT-DEP is not fully consistent, as discussed in appendix B.

7. Experimental setup

The model noise incorporation methods detailed in sections 3 and 6 are benchmarked using "twin experiments," where a "truth" trajectory is generated and subsequently estimated by the ensemble DA systems. As indicated by Eqs. (1) and (2), stochastic noise is added to the truth trajectory and observations, respectively. As defined in Eq. (1), $\mathbf{Q}$ implicitly includes a scaling by the model time step, Δt, which is the duration between successive time indices. Observations are not taken at every time index, but after a duration, Δt_{obs}, called the DA window, which is a multiple of Δt.

The noise realizations excepted, the observation process, Eq. (2), given by $\mathbf{H}$, $\mathbf{R}$, and Δt_{obs}, and the forecast process, Eq. (1), given by f, $\boldsymbol{\mu}^0$, $\mathbf{P}^0$, and $\mathbf{Q}$, are both perfectly known to the DA system. The analysis update is performed using the symmetric square root update of section 2 for all of the methods under comparison. Thus, the only difference between the ensemble DA systems is their model noise incorporation method.

Performance is measured by the root-mean-square error of the ensemble mean, given by

$$\text{RMSE} = \sqrt{\frac{1}{m}\|\bar{\mathbf{x}}^t - \mathbf{x}^t\|_2^2}, \tag{60}$$

for a particular time index t. By convention, the RMSE is measured only immediately following each analysis update. In any case, there was little qualitative difference to "forecast" RMSE averages, which are measured right *before* the analysis update. The score is averaged for all analysis times after an initial transitory period whose duration is estimated beforehand by studying the RMSE time series. Each experiment is repeated 16 times with different initial random seeds. The empirical variances of the RMSEs are checked to ensure satisfying convergence.

TABLE 2. Inflation factors used in benchmark experiments. Reads from left to right, corresponding to the abscissa of the plotted data series.

	Postanalysis inflation									
Fig. 2	None									
Fig. 3	1.25	1.22	1.19	1.15	1.13	1.12	1.10	1.03	1.00	1.00
Fig. 4	1.13	1.25	1.30	1.35	1.43	1.50	1.57	1.65	1.70	
Fig. 5	1.02	1.02	1.02	1.03	1.04	1.05	1.07	1.09	1.13	...
	1.17	1.21	1.31							

Covariance localization is not used. Following each analysis update, the ensemble anomalies are rescaled by a scalar inflation factor intended to compensate for the consequences of sampling error in the analysis (e.g., Anderson and Anderson 1999; Bocquet 2011). This factor, listed in Table 2, was approximately optimally tuned prior to each experiment. In this tuning process the ADD-Q method was used for the forecast noise incorporation, putting it at a slight advantage relative to the other methods.

In addition to the EnKF with different model incorporation methods, the twin experiments are also run with the standard methods of Table 1 for comparison, as well as three further baselines: (i) the climatology, estimated from several long, free runs of the system; (ii) 3D-Var (optimal interpolation) with the background from the climatology; and (iii) the extended Kalman filter (Rodgers 2000).

a. The linear advection model

The linear advection model evolves according to

$$x_i^{t+1} = 0.98x_{i-1}^t, \tag{61}$$

for $t = 0, \ldots,$ $i = 1{:}m$, with $m = 1000$, and periodic boundary conditions. The dissipative factor is there to counteract amplitude growth due to model noise. Direct observations of the truth are taken at $p = 40$ equidistant locations, with $\mathbf{R} = 0.01\mathbf{I}_p$, every fifth time step.

The initial ensemble members, $\{\mathbf{x}_n^0 \,|\, n = 1{:}N\}$, as well as the truth, $\mathbf{x}^0$, are generated as a sum of 25 sinusoids of random amplitude and phase,

$$x_{i,n}^0 = \frac{1}{c_n}\sum_{k=1}^{25} a_n^k \sin(2\pi k[i/m + \varphi_n^k]), \tag{62}$$

where a_n^k and φ_n^k are drawn independently and uniformly from the interval (0, 1) for each n and k, and the normalization constant, c_n, is such that the standard deviation of each $\mathbf{x}_n^0$ is 1. Note that the spatial mean of each realization of Eq. (62) is zero. The model noise is given by

$$\mathbf{Q} = 0.01\,\text{Var}(\mathbf{x}^0). \tag{63}$$

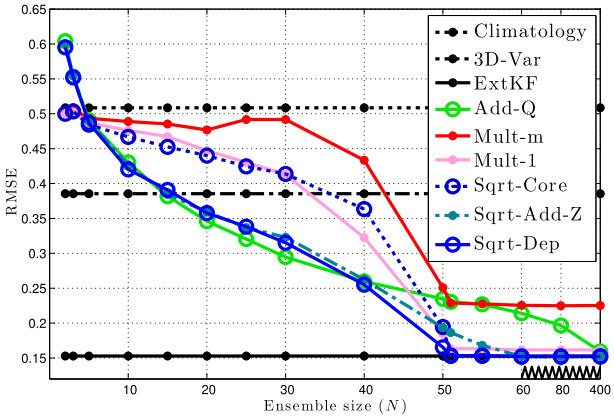

FIG. 2. Performance benchmarks as a function of the ensemble size, N, obtained with the linear advection system. The scale has been irregularly compressed for $N > 60$.

b. *The Lorenz-96 model*

The Lorenz-96 model evolves according to

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \tag{64}$$

for $t > 0$, and $i = 1{:}m$, with periodic boundary conditions. It is a nonlinear, chaotic model that mimics the atmosphere at a certain latitude circle. We use the parameter settings of Lorenz and Emanuel (1998), with a system size of $m = 40$, a forcing of $F = 8$, and the fourth-order Runge–Kutta numerical time-stepping scheme with a time step of $\Delta t = 0.05$. Unless otherwise stated, direct observations of the entire state vector are taken a duration of $\Delta t_{\text{obs}} = 0.05$ apart, with $\mathbf{R} = \mathbf{I}_m$.

The model noise is spatially homogeneous, generated using a Gaussian autocovariance function:

$$\mathbf{Q}_{i,j} = \exp(-1/30\|i-j\|_2^2) + 0.1\delta_{i,j}, \tag{65}$$

where the Kronecker delta, $\delta_{i,j}$, has been added for numerical stability issues.

8. Experimental results

Each figure contains the results from a set of experiments run for a range of some control variable.

a. *Linear advection*

Figure 2 shows the RMSE versus the ensemble size for different model noise incorporation schemes. The maximum wavenumber of Eq. (62) is $k = 25$. Thus, by the design of $\mathbf{P}^0$ and $\mathbf{Q}$, the dynamics will take place in a subspace of rank 50, even though $m = 1000$. This is clearly reflected in the curves of the square root methods, which all converge to the optimal performance of the Kalman filter (0.15) as N approaches 51, and $\mathbf{Z}$ goes to zero. SQRT-ADD-Z takes a little longer to converge because of numerical error. The multiplicative inflation curves are also constant for $N \geq 51$, but they do not achieve the same level of performance. As one would expect, ADD-Q also attains the performance of the Kalman filter for $N \to \infty$.

Interestingly, despite MULT-m satisfying Eq. (29) to a higher degree than MULT-1, the latter performs distinctly better across the whole range of N. This can likely be blamed on the fact that MULT-m has the adverse effect of changing the subspace of the ensemble, though it is unclear why its worst performance occurs near $N = 25$.

ADD-Q clearly outperforms MULT-1 in the intermediate range of N, indicating that the loss of nuance in the covariance matrices of MULT-1 is more harmful than the sampling error incurred by ADD-Q. But, for $45 < N < 400$, MULT-1 beats ADD-Q. It is not clear why this reversal happens.

SQRT-CORE performs quite similar to MULT-1. In the intermediate range, it is clearly deficient compared to the square root methods that account for residual noise, illustrating the importance of doing so. The performance of SQRT-DEP is almost uniformly superior to all of the other methods. The only exception is around $N = 25$, where ADD-Q slightly outperforms it. The computationally cheaper SRA is beaten by ADD for $N < 40$, but has a surprisingly robust performance nevertheless.

b. *Lorenz-96*

Figure 3 shows the RMSE versus ensemble size. As with the linear advection model, the curves of the square root schemes are coincident when $\mathbf{Z} = 0$, which here happens for $N > m = 40$. In contrast to the linear advection system, however, the square root methods still improve as N increases beyond m, and noticeably so until $N = 60$. This is because a larger enable is better able to characterize the non-Gaussianity of the distributions and the nonlinearity of the models. On the other hand, the performance of the multiplicative inflation methods stagnates around $N = m$, and even slightly deteriorates for larger N. This can probably be attributed to the effects observed by Sakov and Oke (2008b).

Unlike the more ambiguous results of the linear advection model, here ADD-Q uniformly beats the multiplicative inflation methods. Again, the importance of accounting for the residual noise is highlighted by the poor performance of SQRT-CORE for $N < 40$. However,

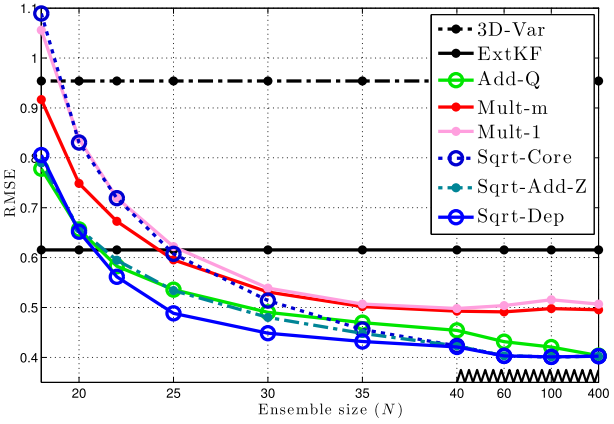

FIG. 3. Performance benchmarks as a function of the ensemble size, N, obtained with the Lorenz-96 system. The climatology averages an RMSE score of 3.7. The scale has been irregularly compressed for $N > 40$.

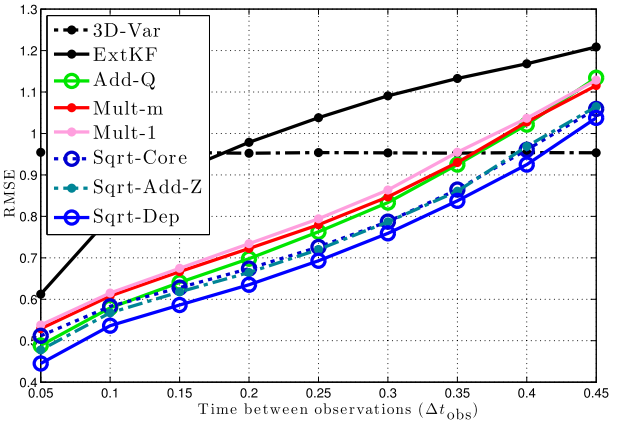

FIG. 4. Performance benchmarks as a function of the data assimilation window, Δt_{obs}, obtained with the Lorenz-96 model and $N = 30$. The climatology averages an RMSE of 3.7.

even though SQRT-ADD-Z is biased, it outperforms ADD-Q for $N > 25$, and approximately equals it for smaller N.

The performance of SQRT-DEP is nearly uniformly the best, the exception being at $N = 18$, where it is marginally beaten by ADD-Q and SQRT-ADD-Z. The existence of this occurrence can probably be attributed to the slight suboptimality discussed in appendix B, as well as the advantage gained by ADD from using it to tune the analysis inflation. Note, though, that this region is hardly interesting, since results lie above the baseline of the extended KF.

ADD-Q asymptotically attains the performance of the square root methods. In fact, though it would have been imperceptible if added to Fig. 3, experiments show that ADD-Q beats SQRT-DEP by an average RMSE difference of 0.005 at $N = 800$, as predicted in section 5.

Figure 4 shows the RMSE versus the DA window. The performance of ADD-Q clearly deteriorates more than that of the deterministic methods as Δt_{obs} increases. Indeed, the curves of SQRT-CORE and ADD-Q cross at $\Delta t_{obs} \approx 0.1$, beyond which SQRT-CORE outperforms ADD-Q. SQRT-CORE even gradually attains the performance of SQRT-ADD-Z, though this happens in a regime where all of the EnKF methods are beaten by 3D-Var. Again, however, SQRT-DEP is uniformly superior, while SQRT-ADD-Z is uniformly the second best. Similar tendencies were observed in experiments (not shown) with $N = 25$.

Figure 5 shows the RMSE versus the amplitude of the noise. Toward the left, the curves converge to the same value as the noise approaches zero. At the higher end of the range, the curves of MULT-m and SQRT-CORE are approximately twice as steep as that of SQRT-DEP. Again, SQRT-DEP performs uniformly superior to the rest, with SQRT-ADD-Z performing second best. In contrasts, ADD-Q performs worse than MULT-m for a noise strength multiplier smaller than 0.2, but better as the noise gets stronger.

9. Summary and discussion

The main effort of this study has been to extend the square root approach of the EnKF analysis step to the forecast step in order to account for model noise. Although the primary motivation is to eliminate the need for simulated, stochastic perturbations, the core method, SQRT-CORE, was also found to possess several other desirable properties, which it shares with the analysis square root update. In particular, a formal survey on these features revealed that the symmetric square root choice for the transform matrix can be beneficial in regards to dynamical consistency.

Yet, since it does not account for the residual noise, SQRT-CORE was found to be deficient in case the noise is strong and the dynamics relatively linear. In dealing with the residual noise, cursory experiments (not shown) suggested that an additive approach works better than a multiplicative approach, similar to the forgetting factor of the SEIK. This is likely a reflection of the relative performances of ADD-Q and MULT-m, as well as the findings of Whitaker and Hamill (2012), which indicate that the additive approach is better suited to account for model error. Therefore, two additive techniques were proposed to complement SQRT-CORE, namely, SQRT-ADD-Z and SQRT-DEP. Adding simulated noise with no components in the ensemble subspace, SQRT-ADD-Z is computationally relatively cheap as well as intuitive. However, it was shown to yield biased covariance

updates due to the presence of cross terms. By re-introducing dependence between the SQRT-CORE update and the sampled, residual noise, SQRT-DEP remedies this deficiency at the cost of an additional SVD.

The utility of the noise integration methods proposed will depend on the properties of the system under consideration. However, SQRT-DEP was found to perform robustly (nearly uniformly) better than all of the other methods. Moreover, the computationally less expensive method SQRT-ADD-Z was also found to have robust performance. These findings are further supported by omitted experiments using fewer observations, larger observation error, and different models.

Future directions

The model noise square root approach has shown significant promise on low-order models, but has not yet been tested on realistic systems. It is also not clear how this approach performs with more realistic forms of model error.

Section c in appendix A shows why it is not possible to eliminate the cross terms, $\mathbf{C}$, which would make SQRT-ADD-Z unbiased. However, there might be a shrewd choice of the square root of $\mathbf{Q}$ that can get close to doing so.

As discussed in appendix B, a more shrewd choice of $\mathbf{Q}^{1/2}$ might improve SRD. This choice impacts $\hat{\boldsymbol{\Xi}}$, but not the core method, as shown in appendix A section c, and should not be confused with the choice of $\mathbf{T}^f$. While the Cholesky factor yielded worse performance than the symmetric choice, other options should be contemplated.

Nakano (2013) proposed a method that is distinct, yet quite similar to SQRT-CORE, this should be explored further, in particular with regards to the residual noise.

Acknowledgments. The authors thank Marc Bocquet for many perspectives and ideas, some of which are considered for ongoing and future research. Additionally, Chris Farmer and Irene Moroz have been very helpful in improving the numerics. The work has been funded by Statoil Petroleum AS, with cofunding by the European FP7 project SANGOMA (Grant 283580).

APPENDIX A

The Residual Noise

a. The cross terms

Let $\mathbf{C}$ be the sum of the two cross terms:

$$\mathbf{C} = \hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}} + \mathbf{Z}\hat{\mathbf{Q}}^{\mathrm{T}/2}, \tag{A1}$$

$$= \boldsymbol{\Pi}_{\mathbf{A}}\mathbf{Q}(\mathbf{I}_m - \boldsymbol{\Pi}_{\mathbf{A}}) + (\mathbf{I}_m - \boldsymbol{\Pi}_{\mathbf{A}})\mathbf{Q}\boldsymbol{\Pi}_{\mathbf{A}}. \tag{A2}$$

Note that $\mathrm{span}(\hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}}) \subseteq \mathrm{span}(\mathbf{A}) \subseteq \ker(\hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}})$, and therefore $\hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}}$ (and its transpose) only has the eigenvalue 0. Alternatively one can show that it is nilpotent of degree 2. By contrast, the nature of the eigenvalues of $\mathbf{C}$ is quite different.

Theorem 3 (properties of $\mathbf{C}$). The symmetry of $\mathbf{C} \in \mathbb{R}^{m^2}$ implies, by the spectral theorem, that its spectrum is real. Suppose that λ is a nonzero eigenvalue of $\mathbf{C}$, with eigenvector $\boldsymbol{v} = \boldsymbol{v}_A + \boldsymbol{v}_B$, where $\boldsymbol{v}_A = \boldsymbol{\Pi}_{\mathbf{A}}\boldsymbol{v}$ and $\boldsymbol{v}_B = (\mathbf{I}_m - \boldsymbol{\Pi}_{\mathbf{A}})\boldsymbol{v}$. Then (i) $\mathbf{u} = \boldsymbol{v}_A - \boldsymbol{v}_B$ is also an eigenvector, (ii) its eigenvalue is $-\lambda$, and (iii) neither $\boldsymbol{v}_A$ nor $\boldsymbol{v}_B$ are zero.

Proof. Note that

$$\mathbf{C}\boldsymbol{v}_A = (\mathbf{I}_m - \boldsymbol{\Pi}_{\mathbf{A}})\mathbf{Q}\boldsymbol{v}_A \in \mathrm{span}(\mathbf{A})^{\perp}, \tag{A3}$$

$$\mathbf{C}\boldsymbol{v}_B = \boldsymbol{\Pi}_{\mathbf{A}}\mathbf{Q}\boldsymbol{v}_B \in \mathrm{span}(\mathbf{A}). \tag{A4}$$

As $\mathbf{C}\boldsymbol{v} = \lambda[\boldsymbol{v}_A + \boldsymbol{v}_B]$, Eqs. (A3) and (A4) imply that

$$\mathbf{C}\boldsymbol{v}_A = \lambda\boldsymbol{v}_B, \tag{A5}$$

$$\mathbf{C}\boldsymbol{v}_B = \lambda\boldsymbol{v}_A. \tag{A6}$$

Therefore,

$$\mathbf{C}\mathbf{u} = \mathbf{C}[\boldsymbol{v}_A - \boldsymbol{v}_B] = \lambda\boldsymbol{v}_B - \lambda\boldsymbol{v}_A = -\lambda[\boldsymbol{v}_A - \boldsymbol{v}_B]. \square$$

Equations (A5) and (A6) can also be seen to imply (iii).

Corollary 1: $\mathrm{trace}(\mathbf{C}) = 0$. This follows from the fact that the trace of a matrix equals the sum of its eigenvalues.

Corollary 2: $\|\boldsymbol{v}_A\|_2^2 = \|\boldsymbol{v}_B\|_2^2$. This follows from the fact that $\boldsymbol{v}^{\mathrm{T}}\mathbf{u} = (\boldsymbol{v}_A + \boldsymbol{v}_B)^{\mathrm{T}}(\boldsymbol{v}_A - \boldsymbol{v}_B) = \boldsymbol{v}_A^{\mathrm{T}}\boldsymbol{v}_A - \boldsymbol{v}_B^{\mathrm{T}}\boldsymbol{v}_B$ should be zero by the spectral theorem.

Interestingly, imaginary, skew-symmetric matrices also have the property that their eigenvalues, all of which are real, come in positive–negative pairs. These matrices can all be written $\mathbf{M} - \mathbf{M}^{\mathrm{T}}$ for some $\mathbf{M} \in i\mathbb{R}^{m^2}$, which is very reminiscent of $\mathbf{C}$. However, it is not clear if these parallels can be used to prove Theorem 3 because $\mathbf{M} - \mathbf{M}^{\mathrm{T}}$ only has zeros on the diagonal, while $\mathbf{C}$ does not (by symmetry, it can be seen that this would imply $\mathbf{C} = 0$). Also, Theorem 3 depends on the fact that the cross terms are "flanked" by orthogonal projection matrices, whereas there are no requirements on $\mathbf{M}$.

b. The residual covariance matrix

The residual, $[\mathbf{Q} - \hat{\mathbf{Q}}]$, differs from the symmetric, positive matrix $\mathbf{Z}\mathbf{Z}^{\mathrm{T}}$ by the cross terms, $\mathbf{C}$. The following theorem establishes a problematic consequence:

Theorem 4 ($[\mathbf{Q} - \hat{\mathbf{Q}}]$ is not a covariance matrix). Provided $\mathbf{C} \neq 0$, the residual "covariance" matrix, $[\mathbf{Q} - \hat{\mathbf{Q}}]$, has negative eigenvalues.

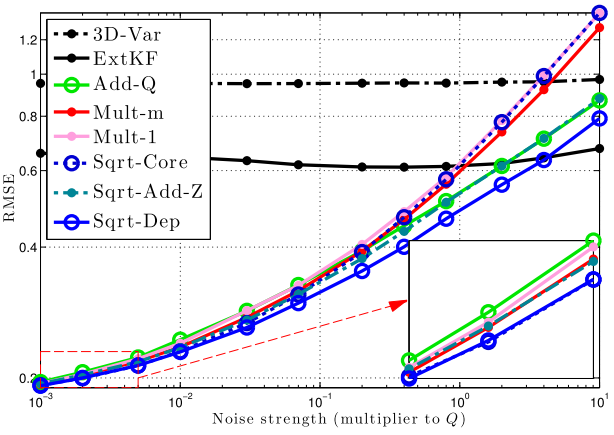

FIG. 5. Performance benchmarks as a function of the noise strength, obtained with the Lorenz-96 model and $N = 25$. Both axes are logarithmic. On average, when $\mathbf{Q}$ is multiplied by 10^{-3} (10^{-2}, 10^{-1}, 10^{0}, 10^{1}, respectively), the model noise makes up approximately 0.5% (4%, 20%, 70%, 90%, respectively) of the growth in the spread of the ensemble between each assimilation. The climatology averages an RMSE score of approximately 4.

Proof. Since $\mathbf{C}$ is symmetric, and thus orthogonally diagonalizable, the assumption that $\mathbf{C} \neq 0$ implies that $\mathbf{C}$ has nonzero eigenvalues. Let $\boldsymbol{v}$ be the eigenvector of a nonzero eigenvalue, and write $\boldsymbol{v} = \boldsymbol{v}_A + \boldsymbol{v}_B$, with $\boldsymbol{v}_A \in \text{span}(\mathbf{A})$ and $\boldsymbol{v}_B \in \text{span}(\mathbf{A})^{\perp}$. Then $\boldsymbol{v}^{\mathrm{T}}\mathbf{C}\boldsymbol{v} = \boldsymbol{v}_A^{\mathrm{T}}\mathbf{Q}\boldsymbol{v}_B \neq 0$. Define $\boldsymbol{v}_\alpha = \boldsymbol{v}_B + \alpha\boldsymbol{v}_A$. Then:

$$\boldsymbol{v}_\alpha^{\mathrm{T}}[\mathbf{Q} - \hat{\mathbf{Q}}]\boldsymbol{v}_\alpha = \boldsymbol{v}_\alpha^{\mathrm{T}}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}} + \mathbf{C}]\boldsymbol{v}_\alpha, \tag{A7}$$

$$= \boldsymbol{v}_B^{\mathrm{T}}\mathbf{Q}\boldsymbol{v}_B + 2\alpha\boldsymbol{v}_A^{\mathrm{T}}\mathbf{Q}\boldsymbol{v}_B. \tag{A8}$$

The second term can always be made negative, but larger in magnitude than the first, simply by choosing the sign of α and making it sufficiently large. □

c. *Eliminating the cross terms*

Can the cross terms be entirely eliminated in some way? Section 6b already answered this question in the negative: there is no particular choice of the square root of $\mathbf{Q}$, inducing a choice of $\hat{\mathbf{Q}}^{1/2}$ and $\mathbf{Z}$ through Eqs. (45) and (46), that eliminates the cross terms: $\mathbf{C} = \hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}} + \mathbf{Z}\hat{\mathbf{Q}}^{\mathrm{T}/2} = 0$.

But suppose we allow changing the ensemble subspace. For example, suppose the partition $\mathbf{Q}^{1/2} = \hat{\mathbf{Q}}^{1/2} + \mathbf{Z}$ uses the projector onto the N largest-eigenvalue eigenvectors of $\mathbf{Q}$ instead of Π_{A}. It can then be shown that the cross terms are eliminated: $\hat{\mathbf{Q}}^{1/2}\mathbf{Z}^{\mathrm{T}} = 0$, and hence $\mathbf{C} = 0$ and $\text{Var}(\mathbf{q}^{\perp\perp}) = \mathbf{Q}$. A similar situation arises in the case of the COFFEE algorithm (see section 5) explaining why it does not have the cross term problem. Another particular rank-N square root, for which $\mathbf{C} = 0$ is the lower-triangular Cholesky factor of $\mathbf{Q}$ with the last $m - N$ columns set to zero.

Unfortunately, for general $\mathbf{Q}$ and $\mathbf{A}$, the ensemble subspace will not be that of the rank-N truncated Cholesky or eigenvalue subspace. Therefore, neither of these options can be carried out using a right-multiplying square root.

APPENDIX B

Consistency of SQRT-DEP

SQRT-CORE ensures that Eq. (30) is satisfied, that is, that

$$\frac{1}{N-1}[\mathbf{A} + \hat{\mathbf{D}}][\mathbf{A} + \hat{\mathbf{D}}]^{\mathrm{T}} = \overline{\mathbf{P}} + \hat{\mathbf{Q}}, \tag{B1}$$

where $(N-1)\overline{\mathbf{P}} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$. However, this does not imply that $\hat{\mathbf{D}}\hat{\mathbf{D}}^{\mathrm{T}} = (N-1)\hat{\mathbf{Q}}$. Therefore, with regards to SQRT-DEP, $\hat{\boldsymbol{\Xi}}\hat{\boldsymbol{\Xi}}^{\mathrm{T}} \neq \mathbf{I}_m$. Instead, the magnitudes of $\hat{\mathbf{D}}$ and $\hat{\boldsymbol{\Xi}}$ are minimized as much as possible, as per Theorem 2.

However, SQRT-DEP is designed assuming that $\hat{\boldsymbol{\Xi}}$ is stochastic, with its columns drawn independently from $\mathcal{N}(0, \mathbf{I}_m)$. If this were the case, then SQRT-DEP would be consistent in the sense of

$$\frac{1}{N-1}\mathbb{E}([\mathbf{A} + \hat{\mathbf{D}} + \tilde{\mathbf{D}}][\mathbf{A} + \hat{\mathbf{D}} + \tilde{\mathbf{D}}]^{\mathrm{T}}) = \overline{\mathbf{P}} + \mathbf{Q}, \tag{B2}$$

where the expectation is with respect to $\tilde{\boldsymbol{\Xi}}$ and $\hat{\boldsymbol{\Xi}}$. This follows from the consistency of $\mathbf{q}$ as defined in Eq. (55), which has $\text{Var}(\mathbf{q}) = \mathbf{Q}$, because each column of $\mathbf{D} = \hat{\mathbf{D}} + \tilde{\mathbf{D}}$ is sampled in the same manner as $\mathbf{q}$.

The fact that $\hat{\mathbf{D}}$ is in fact not stochastic, as SQRT-DEP assumes, but typically of a much smaller magnitude, suggests a few possible venues for future improvement. For example we speculate that inflating $\hat{\boldsymbol{\Xi}}$ by a factor larger than 1, possibly estimated in a similar fashion to Dee (1995). The value of $\hat{\boldsymbol{\Xi}}$ also depends on the choice of square root for $\hat{\mathbf{Q}}^{1/2}$. It may therefore be a good idea to choose $\hat{\mathbf{Q}}^{1/2}$ somewhat randomly, so as to induce more randomness in the square root "noise," $\hat{\boldsymbol{\Xi}}$. One way of doing so is to apply a right-multiplying rotation matrix to $\hat{\mathbf{Q}}^{1/2}$. Cursory experiments indicate that there may be improvements using either of the above two suggestions.

APPENDIX C

Left-Multiplying Formulation of SQRT-CORE

Lemma 1: The row (and column) space of $\mathbf{T}_s^f = (\mathbf{G}^f)_s^{1/2}$ is the row space of $\mathbf{A}$.

Proof. Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$ be the SVD of $\mathbf{A}$. Then:

$$\mathbf{G}^f = \mathbf{I}_N + (N-1)\mathbf{A}^{+}\mathbf{Q}(\mathbf{A}^{+})^{\mathrm{T}},$$

$$= \mathbf{V}[\mathbf{I}_N + (N-1)\boldsymbol{\Sigma}^{+}\mathbf{U}^{\mathrm{T}}\mathbf{Q}\mathbf{U}(\boldsymbol{\Sigma}^{+})^{\mathrm{T}}]\mathbf{V}^{\mathrm{T}}. \square \tag{C1}$$

In view of Lemma 1 it seems reasonable that there should be a left-multiplying update, $\mathbf{A}^f = \mathbf{LA}$, such that it equals the right-multiplying update, $\mathbf{A}^f = \mathbf{AT}^f_s$. Although $N \ll m$ in most applications of the EnKF, the left-multiplying update would be a lot less costly to compute than the right-multiplying one in such cases if $N \gg m$. The following derivation of an explicit formula for $\mathbf{L}$ is very close to that of Sakov and Oke (2008b), except for the addition of Eq. (36). Lemma 2 will also be of use.

Lemma 2. For any matrices, $\mathbf{A} \in \mathbb{R}^{m\times N}$, $\mathbf{M} \in \mathbb{R}^{m^2}$, and any positive integer k,

$$\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{MA})^k = (\mathbf{AA}^{\mathrm{T}}\mathbf{M})^k\mathbf{A}. \tag{C2}$$

Theorem 5 (Left-multiplying transformation): For any ensemble anomaly matrix, $\mathbf{A} \in \mathbb{R}^{m\times N}$, and any SPD matrix $\mathbf{Q} \in \mathbb{R}^{m^2}$,

$$\mathbf{AT}^f_s = \mathbf{LA}, \tag{C3}$$

where

$$\mathbf{T}^f_s = [\mathbf{I}_N + (N-1)\mathbf{A}^+\mathbf{Q}(\mathbf{A}^+)^{\mathrm{T}}]^{1/2}_s, \tag{C4}$$

$$\mathbf{L} = [\mathbf{I}_m + (N-1)\mathbf{AA}^+\mathbf{Q}(\mathbf{AA}^{\mathrm{T}})^+]^{1/2}. \tag{C5}$$

In case $N > m$, Eq. (C5) reduces to

$$\mathbf{L} = [\mathbf{I}_m + (N-1)\mathbf{Q}(\mathbf{AA}^{\mathrm{T}})^{-1}]^{1/2}. \tag{C6}$$

Note that $[\mathbf{I}_m + \mathbf{AA}^+\mathbf{Q}(\mathbf{AA}^{\mathrm{T}})^+]$ is not a symmetric matrix. We can nevertheless define its square root as the square root obtained from its eigendecomposition, as was done for the symmetric square root in section 2c.

Proof. Assuming $\mathbf{A}^+\mathbf{Q}(\mathbf{A}^+)^{\mathrm{T}}$ has eigenvalues less than 1, we can express the square root, $[\mathbf{A}^+\mathbf{Q}(\mathbf{A}^+)^{\mathrm{T}}]^{1/2}$, through its Taylor expansion (Golub and Van Loan 1996, Theorem 9.1.2). Applying Eq. (36), followed by Lemma 2 with $\mathbf{M} = (\mathbf{AA}^{\mathrm{T}})^+(N-1)\mathbf{Q}(\mathbf{AA}^{\mathrm{T}})^+$, and Eq. (36) the other way again, one obtains Eq. (C5).

If $N > m$, then $\mathrm{rank}(\mathbf{A}) = m$, unless the dynamics have made some of the anomalies collinear. Hence, $\mathrm{rank}(\mathbf{AA}^{\mathrm{T}}) = m$ and so $\mathbf{AA}^{\mathrm{T}}$ is invertible, and $\mathbf{AA}^+ = \mathbf{I}_m$. Thus, Eq. (C5) reduces to Eq. (C6). □

Note that the existence of a left-multiplying formulation of the right-multiplying operation $\mathbf{A} \mapsto \mathbf{AT}^f_s$ could be used as a proof for Theorem 1, because $\mathbf{LA}\mathbb{1} = 0$ by the definition of Eq. (28) of $\mathbf{A}$. Finally, Theorem 6 provides an indirect formula for $\mathbf{L}$.

Theorem 6 (Indirect left-multiplying formula): If we have already calculated the right-multiplying transform matrix $\mathbf{T}^f_s$, then the we can obtain a corresponding left-multiplying matrix, $\mathbf{L}$, from

$$\mathbf{L} = \mathbf{AT}^f_s\mathbf{A}^+. \tag{C7}$$

Proof. We need to show that $\mathbf{LA} = \mathbf{AT}^f_s$. Note that $\mathbf{A}^+\mathbf{A}$ is the orthogonal (and hence symmetric) projector onto the row space of $\mathbf{A}$, which Lemma 1 showed is also the row and column space of $\mathbf{T}^f_s$. Therefore, $\mathbf{T}^f_s(\mathbf{A}^+\mathbf{A}) = \mathbf{T}^f_s$, and $\mathbf{LA} = \mathbf{AT}^f_s(\mathbf{A}^+\mathbf{A}) = \mathbf{AT}^f_s$. □

REFERENCES

Anderson, B. D. O., and J. B. Moore, 1979: *Optimal Filtering.* Prentice-Hall, 367 pp.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

——, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, doi:10.1111/j.1600-0870.2008.00361.x.

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758, doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

Arnold, H. M., I. M. Moroz, and T. N. Palmer, 2013: Stochastic parametrizations and model uncertainty in the Lorenz'96 system. *Philos. Trans. Roy. Soc. London*, **A371**, doi:10.1098/rsta.2011.0479.

Ben-Israel, A., and T. N. E. Greville, 2003: *Generalized Inverses: Theory and Applications.* 2nd ed. Springer-Verlag, xvi + 420 pp.

Berry, T., and J. Harlim, 2014: Linear theory for filtering nonlinear multiscale systems with model error. *Proc. Roy. Soc. London*, **A470**, doi:10.1098/rspa.2014.0168.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bocquet, M., 2011: Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes Geophys.*, **18**, 735–750, doi:10.5194/npg-18-735-2011.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Chandrasekar, J., I. S. Kim, D. S. Bernstein, and A. J. Ridley, 2008: Reduced-rank unscented Kalman filtering using Cholesky-based decomposition. *Int. J. Control*, **81**, 1779–1792, doi:10.1080/00207170801891724.

Chen, Y., and D. S. Oliver, 2013: Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.*, **17**, 689–703, doi:10.1007/s10596-013-9351-5.

Daley, R., 1992: Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 1735–1746, doi:10.1175/1520-0493(1992)120<1735:EMECFA>2.0.CO;2.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145, doi:10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2.

Deng, Z., Y. Tang, and H. J. Freeland, 2011: Evaluation of several model error schemes in the EnKF assimilation: Applied to Argo profiles in the Pacific Ocean. *J. Geophys. Res.*, **116**, C09027, doi:10.1029/2011JC006942.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, doi:10.1029/94JC00572.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.

——, 2009: The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Syst.*, **29**, 83–104, doi:10.1109/MCS.2009.932223.

——, and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85–96, doi:10.1175/1520-0493(1996)124<0085:AOGADF>2.0.CO;2.

Golub, G. H., and C. F. Van Loan, 1996: *Matrix Computations.* 3rd ed. Johns Hopkins University Press, 728 pp.

Hamill, T. M., and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.*, **133**, 3132–3147, doi:10.1175/MWR3020.1.

Hanea, R. G., G. J. M. Velders, A. J. Segers, M. Verlaan, and A. W. Heemink, 2007: A hybrid Kalman filter algorithm for large-scale atmospheric chemistry data assimilation. *Mon. Wea. Rev.*, **135**, 140–151, doi:10.1175/MWR3269.1.

Heemink, A. W., M. Verlaan, and A. J. Segers, 2001: Variance reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718–1728, doi:10.1175/1520-0493(2001)129<1718:VREKF>2.0.CO;2.

Hoang, H. S., R. Baraille, and O. Talagrand, 2005: On an adaptive filter for altimetric data assimilation and its application to a primitive equation model, MICOM. *Tellus*, **57A**, 153–170, doi:10.1111/j.1600-0870.2005.00094.x.

Horn, R. A., and C. R. Johnson, 2013: *Matrix Analysis.* 2nd ed. Cambridge University Press, xviii + 643 pp.

Hoteit, I., D.-T. Pham, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *J. Mar. Syst.*, **36**, 101–127, doi:10.1016/S0924-7963(02)00129-X.

——, ——, M. Gharamti, and X. Luo, 2015: Mitigating observation perturbation sampling errors in the stochastic EnKF. *Mon. Wea. Rev.*, **143**, 2918–2936, doi:10.1175/MWR-D-14-00088.1.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, doi:10.1175/MWR-2864.1.

Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273–277, doi:10.1111/j.1600-0870.2004.00066.x.

——, E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, doi:10.1016/j.physd.2006.11.008.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.

Li, H., E. Kalnay, T. Miyoshi, and C. M. Danforth, 2009: Accounting for model errors in ensemble data assimilation. *Mon. Wea. Rev.*, **137**, 3407–3419, doi:10.1175/2009MWR2766.1.

Livings, D. M., S. L. Dance, and N. K. Nichols, 2008: Unbiased ensemble square root filters. *Physica D*, **237**, 1021–1028, doi:10.1016/j.physd.2008.01.005.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, doi:10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108, doi:10.1175/2007MWR2010.1.

Mitchell, L., and A. Carrassi, 2015: Accounting for model error due to unresolved scales within ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **141**, 1417–1428, doi:10.1002/qj.2451.

Nakano, S., 2013: A prediction algorithm with a limited number of particles for state estimation of high-dimensional systems. *16th Int. Conf. on Information Fusion (FUSION)*, Istanbul, Turkey, IEEE,1356–1363.

Nicolis, C., 2004: Dynamics of model error: The role of unresolved scales revisited. *J. Atmos. Sci.*, **61**, 1740–1753, doi:10.1175/1520-0469(2004)061<1740:DOMETR>2.0.CO;2.

Oliver, D. S., 2014: Minimization for conditional simulation: Relationship to optimal transport. *J. Comput. Phys.*, **265**, 1–15, doi:10.1016/j.jcp.2014.01.048.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428, doi:10.1111/j.1600-0870.2004.00076.x.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207, doi:10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2.

Reich, S., and C. J. Cotter, 2013: Ensemble filter techniques for intermittent data assimilation. *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences*, M. Cullen et al., Eds., Radon Series on Computational and Applied Mathematics, Vol. 13, De Gruyter, 91–134.

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice.* World Scientific, 240 pp.

Sakov, P., and P. R. Oke, 2008a: A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus*, **60A**, 361–371, doi:10.1111/j.1600-0870.2007.00299.x.

——, and ——, 2008b: Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Wea. Rev.*, **136**, 1042–1053, doi:10.1175/2007MWR2021.1.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, doi:10.1256/qj.04.106.

Slingo, J., and T. Palmer, 2011: Uncertainty in weather and climate prediction. *Philos. Trans. Roy. Soc. London*, **A369**, 4751–4767, doi:10.1098/rsta.2011.0161.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114, doi:10.1175/2009MWR2835.1.

Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrol. Hydraul.*, **11**, 349–368, doi:10.1007/BF02427924.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast

schemes. *J. Atmos. Sci.*, **60**, 1140–1158, doi:10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

——, ——, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590–1605, doi:10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.

——, G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190–1200, doi:10.1175/1520-0493(2004)132<1190:RWRUED>2.0.CO;2.

——, T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463–482, doi:10.1175/2007MWR2018.1.

Wunsch, C., 2006: *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications.* Cambridge University Press, 384 pp.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **132**, 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zupanski, D., and M. Zupanski, 2006: Model error estimation employing an ensemble data assimilation approach. *Mon. Wea. Rev.*, **134**, 1337–1354, doi:10.1175/MWR3125.1.